

User-Uncertainty: A Human-Centred Uncertainty Taxonomy for VGI through the Visual Analytics Workflow

Alexandra Diehl^{*}, Bin Yang[†], Rahul Deb Das[‡], Siming Chen[§],
Gennady Andrienko[¶], Natalia Andrienko[¶], Doris Dransch[†] and Daniel Keim^{*}

^{*}University of Konstanz, Germany

Email: {alexandra.diehl,keim}@uni-konstanz.de

[†]GFZ Potsdam, Germany

Email: bin.yang@gfz-potsdam.de

[‡]University of Zurich, Switzerland

Email: rahul.das@geo.uzh.ch

[§]University of Bonn / Fraunhofer IAIS, Germany

Email: siming.chen@iais.fraunhofer.de

[¶]Fraunhofer IAIS / City, University of London, Germany / UK

Email: {gennady.andrienko,natalia.andrienko}@iais.fraunhofer.de

Abstract—The emergence of Web 2.0 and ubiquitous mobile platforms makes it possible to collect a vast amount of information contributed by people (VGI). For example, crowd-sourcing applications collect information from domains such as biodiversity, urban planning, and risk management, and other sources such as social media connect citizens that exchange voluntarily huge amount of posts on platforms like Twitter, Flickr, and Facebook. VGI differs from data coming from sensors, simulations, and mathematical models. It is highly dependent on the human wills to share the information, and the background and knowledge of the user, which introduces uncertainty. In this paper, we explore different dimensions of VGI uncertainty from the perspective of the human that contributes with the data, as well as the technology and systems used to collect the data. Our contributions include a new taxonomy that explicitly differentiates among the uncertainty introduced by the humans, we named *User-Uncertainty* and analyzed it at different steps of the Visual Analytics Workflow, and several use cases that illustrate our approach for the case of User-Uncertainty coming from the producers. We conclude our paper with a discussion about the potential uses and future work to be done to understand User-Uncertainty.

Index Terms—VGI, Visual Analytics, Uncertainty

I. INTRODUCTION

With the emergence of Web 2.0 and ubiquitous mobile platforms people can contribute vast amount of information, voluntarily, in space and time to different platforms and purposes. This information contributed by people can be framed into the term volunteered-geographic information (VGI). Goodchild coined the term VGI [5] as geographic information that is provided voluntarily, that presents inaccuracies and it is mostly provided by untrained people. VGI has become an important source of information for many stakeholders in different contexts [5], [12]. It feeds systems to model human behavior,

environmental phenomena, species distribution models, and it is used in risk management and decision-making, as many of other application areas. The analysis of the VGI uncertainty is essential to increase the users' trustability on the models and systems. Visual Analytics has been preached for almost two decades the need of including "the human in the loop" to make sense of the data, the intermediate process, and final results. Paradoxically, to include "the user in the loop" to optimize the analytic process also implies to tackle the uncertainty introduced by the users as part of the human reasoning and decision making.

VGI differs from data coming from sensors, simulations, and mathematical models because it is highly dependent on the human. Humans are the producers and the consumers of VGI. The "human factor" introduces a new type of uncertainty into the analytic workflow. This uncertainty can be found when the VGI is input into the system, at each intermediate step of the visual analytic process, and at the very last step, when the results of the analytic process are communicated through reports and visual abstractions. Our goal is to analyze the humans factors of the uncertainty in VGI, which we enclosed into the term User-Uncertainty. We also study its interrelation with the spatial, temporal, and thematic uncertainty types introduced in previous literature.

The study of User-Uncertainty poses some challenging questions. Among them:

- How the User-Uncertainty is embedded in the contributed VGI?
- How the communication of User-Uncertainty can help stakeholders and decision makers to improve their tasks workflow?

- What are the plausible qualitative and quantitative methods to analyze the User-Uncertainty through the process?

In this paper, we define User-Uncertainty as a new type of uncertainty coming from the conceptualizations, actions, and decision making of VGI users. We propose a new taxonomy that includes User-Uncertainty, and discuss its interrelations with spatial, temporal, and thematic uncertainty. We conclude our paper with a discussion of our new classification in the light of the above-mentioned research questions.

II. RELATED WORK

We divide the related work into two main aspects: (1) a review of current visual analytics models of uncertainty, and (2) a review of different existing taxonomies and classifications of VGI Uncertainty.

A. Visual Analytics Models of Uncertainty for VGI

During the last years, the study of the uncertainty and its propagation through the visual analytics workflow has gained popularity. Dasgupta et al. [4] analyzed the visual uncertainty in visual representations to construct more efficient visualizations. They distinguished between data-uncertainty and visual uncertainty and analyzed points of intersection, for example, when the geometric abstraction is considered as part of visual uncertainty but could also be considered as part of the data-uncertainty. Early, in 2015, MacEachren [9] proposed to consider the propagation of uncertainty through the whole VA workflow rather than just the visualization of the uncertainty at the end of the pipeline. He illustrated current challenges and possible approaches to tackle uncertainty using definitions from decision sciences. In this work, we take a similar approach but we analyze uncertainty at a deeper level, also considering the interrelationship between data-uncertainty and User-Uncertainty. Kinkeldey et al. [8] analyzed the impact of visually represented geodata uncertainty on decision-making, addressed possible approaches for the evaluation of uncertainty in visualizations [7]. Sacha et al. [12] also presented a knowledge generation model that considers the uncertainty of the data and the propagation of uncertainties through the visual analytics workflow. In this work, the authors associated the term uncertainty and accuracy to the data processing and trustworthiness to humans perceptual and cognitive biases. Chen et al. [2] described the concept of "soft knowledge" as information coming from analysts' intuition, theories and beliefs. The uncertainty could come from incomplete, noisy, or contradictory data. It can also be introduced by the user in form of soft knowledge and affected by prior information (previous observations, experimentation, analytic conclusions). The authors proposed a theoretical framework based on information theory to characterize the propagation of knowledge through the data intelligence pipeline.

B. Uncertainty Classification

Senaratne et al [13] provided a comprehensive survey of existing quality measures that are used as uncertainty indicators

for VGI. The authors collected papers from the state-of-the-art where those measures were presented or used in different kind of information, as text-based, image-based, and map-based types. Although previous work mentioned the humans factors of VGI, they do not explicitly separated them as a different concept. We propose a taxonomy that make explicit the humans factors in account of uncertainty. Therefore, we investigated how humans, in particular as producers and consumers of information, deal with uncertainty. There is a vast literature about the relationship between the uncertainty perceived by the user and decision-making [1], [11], some of them, from the visual analytics community [7]–[9]. Tannert et al. [14] proposed a taxonomy of uncertainty from the point of view of decision making, where the user is confronted to specific events. They analyzed uncertainty related to its impacts to decision making, risks, and dangers. In our approach, we provide both a classification of the uncertainty coming from the characteristics of the data (data-uncertainty) and the uncertainty inherent to the user that create, analyze, and communicate the information (user-uncertainty).

III. UNCERTAINTY TAXONOMY

We propose a taxonomy that includes both data-uncertainty and user-uncertainty (see figure 1). We include into data-uncertainty the kind of uncertainty that is inherent to data management and the technology used to collect the data. Examples of data-uncertainty are related to the precision of the instruments for gathering the data, or the fusion of data sources with different resolutions, or from inconsistencies among heterogeneous data sources. We use the state-of-the-art classification of the data-uncertainty based on the data type and the ISO quality standards [6]: spatial-, temporal-, thematic-uncertainty, completeness, and consistency.

We define *User-Uncertainty* as the human uncertainty inherent in VGI, that is introduced into the visual analytics workflow every time the user interacts with the system. User-Uncertainty comes from the way humans give meaning or conceptualize facts, or by the way they action to a specific event. When dealing with VGI, the human conceptualizations and decision-making influence the uncertainty through the complete visual analytics workflow, the human is in the loop and therefore, the uncertainty is in the loop as well.

A. Data-Uncertainty

We address our taxonomy from the decision making point of view because it fits very well to the visual analytics workflow. The user has to reason about facts and make decisions all along the workflow, producing and consuming VGI. For example, when he posts a tweet, the producer makes a decision about what to post and how much to disclose about what he knows.

The top-level of our hierarchy divides uncertainty into data-uncertainty and user-uncertainty. We divide the data-uncertainty branch into explicit and implicit uncertainty similar as described in [3]. Both of them, can be sub-classified as temporal, spatial, and thematic uncertainty. The sources of this kind of uncertainty can be for example: loss of information,

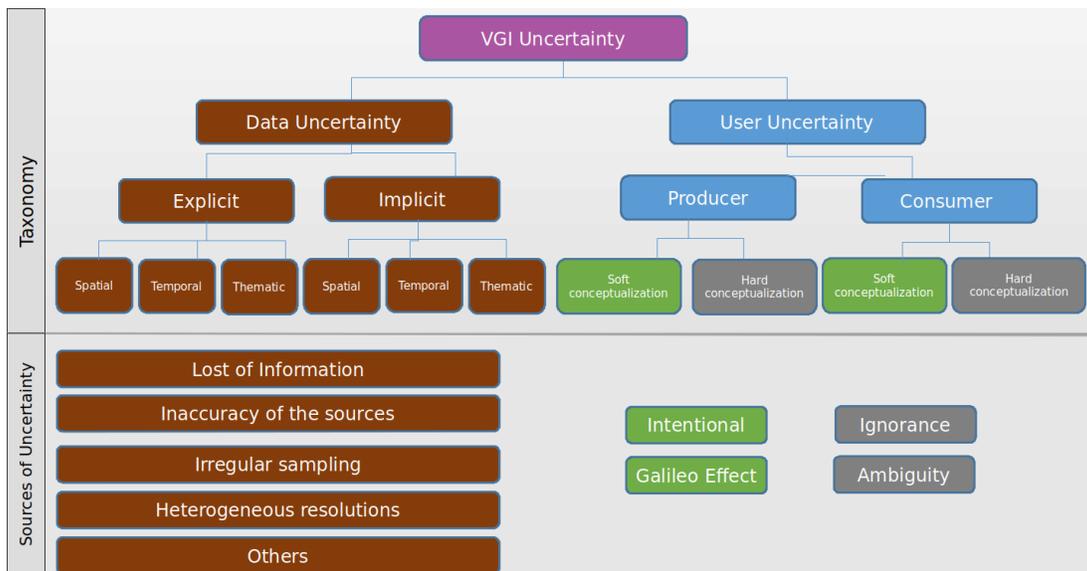


Fig. 1. VGI Uncertainty Ontology of data-uncertainty and user-uncertainty.

inaccuracy of the sources, inconsistency, irregular sampling, etc.

B. User-Uncertainty

The **producers** are the volunteers that input the data into the system. The **consumers** are the analysts or decision makers that work with the data to solve specific problems.

We divide the user-uncertainty branch depending on the type of human conceptualization: **soft-conceptualization** and **hard-conceptualization**. The conceptualization, as "an abstract simplified view of some selected part of the world" [15] can be affected in different ways. We have classified them into four main categories: **intentional**, by **ignorance**, by a strong belief or bias named in literature as "**Galileo Effect**" [14], and by **ambiguity** as the state where there is no scientific method to determine one or another direction to go.

C. Uncertainty in the Visual Analytics Loop

We illustrate our taxonomy ontology in each step of the visual analytics workflow, according to spatial, temporal, thematic, and entity of VGI features (Fig. 2). In each step of the workflow, there can be user-uncertainty and data-uncertainty. For example, in spatial analysis, the processing techniques may lead to spatial detail loss, e.g. using statistic results of spatial distribution. In such case, the data-uncertainty is introduced. In the other perspective, users could make mistakes, misjudge or ignore some parts of the spatial distribution, which involves user-uncertainty. Such uncertainty should be taken into account because of the user's role in the loop. More examples can be referred from the proposed ontology in each visual analytics process.

IV. APPLICABILITY

To illustrate the applicability of our taxonomy we present several real examples of traffic-related tweets in the context

of a highly populated city. These examples focus only on the user as a *producer* of VGI. Further investigation needs to be done on visual assessment systems that can help us to understand the user-uncertainty introduced by the consumer. The following scenarios illustrate how the user-uncertainty relates to the thematic, spatial, and temporal aspects of data-uncertainty.

We focus our case study on the expressions of uncertainty contained on the tweets. In this scenarios, we consider that people may express their uncertainty through some action. We used verb or adjective form of different conceptualization aspects (e.g., ignorance → ignore; confusion → confused; completeness → complete; belief → believe). We use the keyword-based approach shown in figure 3 to analyze the tweets. To develop our lexicon for the user-uncertainty, we leverage a graph-based lexical dictionary which models the semantic relationship between different words. We use Wordnet [10] to create the synsets clusters of synonyms. It is plausible that a given word may appear in different synsets based on its semantics.

The following scenarios tackle the first research question: (1) "How the user-uncertainty introduced by the producers is embedded in the information?". We collected some tweets tweeted by people in a highly populated city, related to traffic events in the city. We analyze each tweet considering user-uncertainty, and spatial, temporal and thematic uncertainty. The names of the roads and cities were anonymized.

Tweet 1 . Daily heavy **traffic** in anonymized road,, no support from govt, wasting hell lot of fuel daily.... Govt **ignore**....

In this tweet, the producer contributed traffic information on a daily basis. The spatial scope is the anonymized road. However, it is not mentioned where exactly traffic takes place on that road. In this tweet, the producer also expressed his

VA Process VGI Features	Data Acquisition	Data Processing	Data Analysis	Data Visualization
Spatial	[D] GPS Error Range, Coordinates Level Granularity	[D] Aggregation Information Loss	[D] Spatial Detail Loss (Statistics Results of Spatial distribution)	[D] Visual Clutter
	[U] Intentional Error	[U] Geo Inference Accuracy	[U] Mistakes, Ignorance	[U] Visual spatial perception accuracy, Spatial Interpolation Estimation
Temporal	[D] Time Recording Error Range, Coordinates Level Granularity Time Zone Information Loss	[D] Trending Prediction Errors	[D] Time Prediction Accuracy	[D,U] Temporal Interpolation Estimation
		[U] Time Inference Accuracy		
Thematic	[D] Recording loss	[D] Information Loss (Stop Words Removing, Extract top words, etc.)	[D] Probability description (Topic Extraction)	[D] Information Loss (Top words)
	[U] Typo, Ambiguity	[U] Galileo Effect	[U] Sentimental Estimation Bias	[U] Mis-interpretation, Semantic Guessing
Entity	[U] Fake information, Intentional Error	[U] Inaccurate summary / profiling, Ignorance	[D, U] Categorization Correctness	[U] Perceptual Errors
			[U] Analysis Perspective	

Fig. 2. VGI Uncertainty Ontology with illustrative examples for a visual analytics Workflow, including data acquisition (producer), data processing (consumer), data analysis (consumer) and data visualization (consumer).

Hard conceptualization		
Conceptualization aspect	Elements in the synset	Synset definition by Wordnet
Ignorance (ignore)	'ignore'	fail to notice
	'neglect', 'ignore', 'disregard', 'ignore'	give little or no attention to be ignorant of or in the dark about
Confusion (confused)	'confuse', 'confound'	mistake one thing for another
	'confuse', 'throw', 'fox', 'befuddle', 'fuddle', 'bedevil', 'confound', 'discombobulate'	be confusing or perplexing to; cause to be unable to think clearly
	'confuse', 'flurry', 'disconcert', 'put_off'	cause to feel embarrassment
	'baffled', 'befuddled', 'bemused', 'bewildered', 'confounded', 'confused', 'lost', 'mazed', 'mixed-up', 'at_sea'	perplexed by many conflicting situations or statements; filled with bewilderment
	'confused', 'disconnected', 'disjointed', 'disordered', 'garbled', 'illogical', 'scattered', 'unconnected'	lacking orderly continuity
Soft Conceptualization		
Intentional (intent)	'purpose', 'intent', 'intention', 'aim', 'design'	an anticipated outcome that is intended or that guides your planned actions
	'intent', 'purport', 'spirit'	the intended meaning of a communication
	'captive', 'absorbed', 'engrossed', 'enwrapped', 'intent', 'wrapped'	giving or marked by complete attention to
Strong belief that causes bias / "Galileo Effect"	'believe'	accept as true; take to be true
	'think', 'believe', 'consider', 'conceive'	judge or regard; look upon; judge
	'believe', 'trust'	be confident about something
	'ignore', 'disregard', 'snub', 'cut'	refuse to acknowledge
	'dismiss', 'disregard', 'brush_aside', 'brush_off', 'discount', 'push_aside', 'ignore'	bar from attention or consideration

Fig. 3. Different clusters of synset elements and their given semantics in the context of the new proposed taxonomy.

conceptualization of ignorance of the facts by the authority.

Tweet 2. *I agree. What is happening in anonymized-city is that cops are actually being asked to ignore traffic offences and manage traffic instead. The idea is that the cameras will do the job. But again, the sheer volumes...*

The above tweet reflects the understanding of the producer regarding the current traffic condition in the city. The tweet shows the (implicit) belief of the producer that authority has instructed the cops to overlook traffic offence. This could be associated with the "strong belief" uncertainty.

Tweet 3. *Railway Derailment Hatrick by anonymized Govt anonymized-CST Harbor Local Train Derailed in anonymized-city Coz Gross Neglect of Aging Infrastructure*

In this tweet, the producer informed about an accident in anonymized-city, however, with a very high spatial uncertainty as it is not yet understood where the accident took place. The information also involves high temporal uncertainty. The producer also expressed strong belief of authority negligence to the transport infrastructure.

Tweet 4. *The car parked right in the middle of the road anonymized near anonymized Stn next to anonymized shopping centre this is causing traffic jams.*

It is also observed that producers provide evidence (by the fact, picture or video or web link) of action and express their pragmatic belief. In this example, the producer reported about a traffic jam and illegitimate parking as the reason. However, while reporting the location of illegitimate parking, the producer tried to be specific and used a spatial cue, "near", to handle spatial uncertainty in an explicit way.

V. CONCLUSIONS

These examples above show how producers can introduce uncertainty based on their understanding or state of belief. More research has to be done to understand how producers can introduce uncertainty by omission. A possible approach could be to use the collective behavior of users as a ground truth to compare with. We only illustrated our concepts using textual data, but the same uncertainties can also be described with other VGI data types e.g., media data (Flickr, Instagram), map data (OpenStreetMap).

In the upcoming work we will continue the development of our ontological model to leverage the data-uncertainty and User-Uncertainty through the complete visual analytics process. We will also tackle the open questions stated in the introduction about the communication of the User-Uncertainty to help stakeholders and decision makers to improve their tasks workflow, and plausible qualitative and quantitative methods to analyze the User-Uncertainty through the process.

ACKNOWLEDGMENT

We would like to thank our colleagues Sanae Mahtal, Rene Westerholt, and Franz-Benjamin Mocnik for the inspiring discussions.

REFERENCES

- [1] A. Akhunzada, M. Sookhak, N. B. Anuar, A. Gani, E. Ahmed, M. Shiraz, S. Furnell, A. Hayat, and M. K. Khan. Man-at-the-end attacks: Analysis, taxonomy, human aspects, motivation and future directions. *Journal of Network and Computer Applications*, 48:44–57, 2015.
- [2] M. Chen and A. Golan. What may visualization processes optimize? *IEEE transactions on visualization and computer graphics*, 22(12):2619–2632, 2016.
- [3] M. Craglia, F. Ostermann, and L. Spinsanti. Digital earth from vision to practice: making sense of citizen-generated content. *International Journal of Digital Earth*, 5(5):398–416, 2012.
- [4] A. Dasgupta, M. Chen, and R. Kosara. Conceptualizing visual uncertainty in parallel coordinates. In *Computer Graphics Forum*, volume 31, pages 1015–1024. Wiley Online Library, 2012.
- [5] M. F. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.
- [6] J. B. Guinée. Handbook on life cycle assessment operational guide to the ISO standards. *The international journal of life cycle assessment*, 7(5):311, 2002.
- [7] C. Kinkeldey, A. M. MacEachren, M. Riveiro, and J. Schiewe. Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science*, 44(1):1–21, 2017.
- [8] C. Kinkeldey, A. M. MacEachren, and J. Schiewe. How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4):372–386, 2014.
- [9] A. M. MacEachren. Visual analytics and uncertainty: Its not about the data. 2015.
- [10] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [11] W. D. Rowe. Understanding uncertainty. *Risk analysis*, 14(5):743–750, 1994.
- [12] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The role of uncertainty, awareness, and trust in visual analytics. *IEEE transactions on visualization and computer graphics*, 22(1):240–249, 2016.
- [13] H. Senaratne, A. Mobasheri, A. L. Ali, C. Capineri, and M. Haklay. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1):139–167, 2017.
- [14] C. Tannert, H.-D. Elvers, and B. Jandrig. The ethics of uncertainty: In the light of possible dangers, research becomes a moral duty. *EMBO reports*, 8(10):892–896, 2007.
- [15] Wikipedia contributors. Conceptualization (information science) — Wikipedia, the free encyclopedia, 2017. [Online; accessed 22-August-2018].